nature ecology & evolution

Article

https://doi.org/10.1038/s41559-024-02587-2

# Genomics of a sexually selected sperm ornament and female preference in *Drosophila*

Received: 19 February 2024

Accepted: 22 October 2024

Published online: 22 November 2024

Check for updates

Zeeshan A. Syed[1] ✉, R. Antonio Gomez[1], Kirill Borziak[1], Amaar Asif[1], Abelard S. Cong[1], Patrick. M. O'Grady[2], Bernard Y. Kim[3], Anton Suvorov[4], Dmitri A. Petrov[3], Stefan Lüpold[5], Peter Wengert[1], Caitlin McDonough-Goldstein[1], Yasir H. Ahmed-Braimah[1], Steve Dorus[1] ✉ & Scott Pitnick[1] ✉

Our understanding of animal ornaments and the mating preferences driving their exaggeration is limited by knowledge of their genetics. Post-copulatory sexual selection is credited with the rapid evolution of female sperm-storage organ morphology and corresponding sperm quality traits across diverse taxa. In *Drosophila*, the mechanisms by which longer flagella convey an advantage in the competition among sperm for limited storage space in the female, and by which female sperm-storage organ morphology biases fertilization in favour of longer sperm have been resolved. However, the evolutionary genetics underlying this model post-copulatory ornament and preference system have remained elusive. Here we combined comparative analyses of 149 *Drosophila* species, a genome-wide association study in *Drosophila melanogaster* and molecular evolutionary analysis of ~9,400 genes to elucidate how sperm and female sperm-storage organ length co-evolved into one of nature's most extreme ornaments and preferences. Our results reveal a diverse repertoire of pleiotropic genes linking sperm length and seminal receptacle length expression to central nervous system development and sensory biology. Sperm length development appears condition-dependent and is governed by conserved hormonal (insulin/insulin-like growth factor) and developmental (including *Notch* and *Fruitless*) pathways. Central developmental pathway genes, including *Notch*, also comprised the majority of a restricted set of genes contributing to both intraspecific and interspecific variation in sperm length. Our findings support 'good genes' models of female preference evolution.

The theory of sexual selection has been a triumph in the quest to explain the origin and maintenance of biodiversity among sexually reproducing organisms[1,2]. Intrasexual competition for matings (more prominent between males) and intersexual choice among potential mates (more prominent by females) have contributed to striking sexual dimorphism in species and variation among species in sexual armaments and ornaments[1,2]. Because sexual selection continues after mating whenever females are polyandrous[3,4], post-copulatory sexual selection is similarly credited with the diversification of genitalia, sperm, seminal fluid and female reproductive tract structure and function[4–7].

Resolving the complex genetic basis of female preferences and male sexual structures, displays and signals is critical for identifying the developmental basis of their expression and diversification, understanding the maintenance of genetic variation and discriminating among alternative models for female preference evolution[8,9]. Critical tests of sexual selection theory have come, for example, from studies

A full list of affiliations appears at the end of the paper. ✉e-mail: zeeshyeda@gmail.com; sdorus@syr.edu; sspitnic@syr.edu

revealing that: (1) a heterozygote advantage at a single locus, mediated by a trade-off between reproductive success and survival, maintains genetic variation in horn size of wild male Soay sheep[10]; (2) sex-specific regulation of a HOX gene underlies convergent evolution of male sex combs in *Drosophila*[11]; (3) experimentally intensifying sexual selection on a weapon in a bulb mite increases genome-wide purifying selection[12]; and (4) condition-dependence of horn growth in dung beetles is attributable to their sensitivity to signalling through the insulin/insulin-like growth factor (IGF) pathway[13]. Nevertheless, the endeavour of resolving the genomics of sexual selection has proven difficult because of the challenges of investigating indirect genetic effects and pleiotropy, and of applying 'omic' technologies to taxa that have traditionally served as models for the study of sexual selection[2,9]. Perhaps an even greater obstacle is quantifying variation in female mate preferences, which entail the perception of signals from potential mates, central nervous system (CNS) processing of those perceptions and subsequent behavioural responses by the female[8].

Here we surmount many of these obstacles through genomic analyses of a post-copulatory preference and ornament in *Drosophila*[5,14]. Examination of the evolutionary allometry of exaggerated male sexual traits throughout the animal kingdom, including pheasant tail feathers, cervid antlers, lizard dewlaps and scarab beetle horns, revealed that sperm flagellum length in *Drosophila* is one of nature's most extreme ornaments[15]. It is more variable among members of the family Drosophilidae than among the remainder of the animal kingdom[16] (range: 0.19 ± 0.00 mm in *Scaptodrosophila latifasciaeformis* to 58.29 ± 0.66 mm in *Drosophila bifurca* for species in the current study) (Supplementary Data 1). Longer sperm are relatively costly to manufacture[15,17], but are better at displacing and resisting displacement by shorter sperm when competing for space in the female's primary sperm-storage organ: the seminal receptacle (SR)[18–20]. The extent of these competitive fertilization advantages increases with increasing SR length (which also engenders substantive developmental and maintenance costs[21]) and is mediated by interactions between the sexes[18–20]. Hence, the strength of female preference for longer sperm in *Drosophila* can be reliably quantified simply by measuring SR length (range: 0.22 ± 0.01 mm in *S. latifasciaeformis* to 81.65 ± 2.39 mm in *D. bifurca* for species in the current study) (Supplementary Data 1).

We utilized this post-copulatory ornament (sperm length) and female preference (SR length) in *Drosophila* to address several long-standing questions in evolutionary biology. (1) Have the female preference and male ornament co-evolved (that is, each trait has been part of the selective environment of the reciprocal trait), and how does variation in preference strength relate to rate of ornament diversification? (2) What are ornament genes and how do they align with theoretical models of female preference evolution? (3) Do conserved developmental pathways underlie ornament elaboration? (4) Is the recent selective history of an ornament consistent with the 'lek paradox'?[22,23] (5) Do the same genes underlying intraspecific variation in a trait contribute to that trait's extreme variation across species? Here we combine phylogenetic analysis of co-variation between SR and sperm length across 149 *Drosophila* (and related genera) species that have greatly diversified over ~65 million years, a genome-wide association study (GWAS) of SR length and sperm length in *Drosophila melanogaster* and molecular evolutionary analysis of ~9,400 genes relative to SR length and sperm length to examine the genomics of sexual selection.

## Results and discussion

### The female preference and sperm ornament have co-evolved

Sexual selection theory predicts the co-evolution of female preferences and corresponding male ornaments. We determined mean SR length and sperm length for 149 species of drosophilids and then reconstructed the evolutionary history of both traits using a robust time-calibrated phylogeny derived from whole-genome sequence data for all species (Supplementary Figs. 1 and 2 and Supplementary Data 1). Phylogenetic generalized least-squares (PGLS) analysis confirmed[24] a strong pattern of preference–ornament co-diversification (adjusted $R^2 = 0.87$, $P < 2^{-16}$) (Fig. 1a). We next examined the degree to which transitions between discrete sex-specific character states (that is, short versus long sperm and SRs, respectively) across the phylogeny were dependent upon trait expression in the opposite sex in a Bayesian probabilistic framework (Fig. 1b). Models predicting transitions in one sex in response to the other sex found much stronger support than trait-independent models (log(Bayes factor) > 50 for trait-dependent models). Specifically, transition rates between character states were high for changes in sperm length in response to SR length, and moderate for the reverse, supporting preference–ornament co-evolution (Fig. 1b and Supplementary Fig. 3).

We further explored the interdependence of SR length and sperm length diversification by capitalizing on variation among species in the mode of female sperm storage. In addition to the SR, *Drosophila* females also have a pair of spermathecae (SP), which are spheroid capsules at the end of ducts that store sperm in some species. We characterized organ usage for all 149 species and found that: (1) sperm storage in both organs ('SP + SR' species; $n = 107$) is probably the ancestral character state for the tree (posterior probability = 0.61) (Fig. 1c); (2) there have been multiple, independent losses of sperm storage in the SP ('SR only' species; $n = 42$); and (3) nested in those branches are several reversals back to storing sperm in both organ types[24] (Fig. 1c and Supplementary Fig. 4). Summarizing transitions across stochastic mappings revealed that the tree spent most of its time in the 'SP + SR' state (75.71% of the time). Optimal transition probabilities support the existence of strong asymmetries in sperm-storage mode evolution with an average of 3.89 gains of 'SR only' compared with an outsized 13.47 transitions in the reverse direction. Given compelling experimental evidence that only the SR generates selection on sperm length[18,20], we tested the a priori prediction that rate of sperm length evolution should be faster among the 'SR only' species. Models revealed that sperm length evolved approximately 18 times faster, on average, among those species compared with species using both organs (Fig. 1d). Further, trait-dependent modelling of SR length and sperm-storage mode revealed that SR length evolution is positively associated with transitions to the 'SR only' mode (log(Bayes factors) = 7.5), which suggests that post-copulatory sexual selection has favoured the evolutionary loss of use of the SP as sperm-storage organs.

### Sperm ornamentation and female preference genes

We examined intraspecific variation in sperm length and SR length in *D. melanogaster* based on measures of 7,560 sperm from 1,512 males and SRs from 2,016 females across 126 lines of the *Drosophila* Genetic Reference Panel (DGRP), a community resource consisting of fully sequenced inbred lines derived from a single natural population of *D. melanogaster*[25]. We identified considerable among-line variation in both sperm length (coefficient of variation ($C_V$) = 6.96) and SR length ($C_V = 9.75$) (Fig. 2a and Supplementary Data 2) and nearly identical broad-sense heritabilities (sperm length $H^2 = 0.69$; SR length $H^2 = 0.67$). GWAS identified 217 and 54 variants that were significantly associated with sperm length and SR length ($P < 10^{-5}$), respectively (Fig. 2b and Supplementary Data 3). Based on their genomic location, these variants identified 142 candidate sperm length genes and 19 SR length genes. The numeric disparity of variants associated with sperm length and SR length, despite almost identical $H^2$ values, is probably attributable to differences in polygenicity, variant effect sizes and measurement accuracy between the two traits (Methods).

Sexual selection theory (that is, both Fisherian runaway and good genes models) predicts that female choice will result in the accumulation of linkage disequilibrium between preference and ornament genes, which in turn will contribute to their co-diversification[23]. We examined the genetic architecture of the traits and found that variants

https://doi.org/10.1038/s41559-024-02587-2

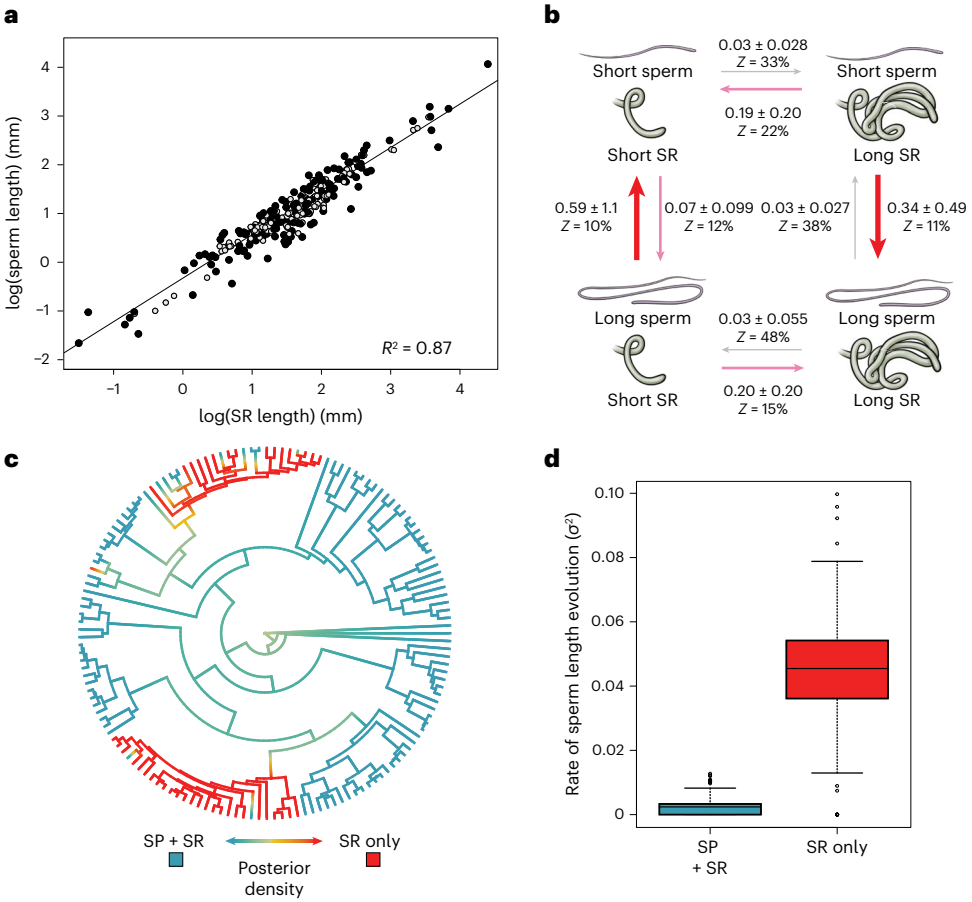


**Fig. 1 | Sperm length and female SR length have co-evolved across the drosophilid tree of life. a**, Relationship between log(sperm length) and log(SR length) for 149 extant species (black) from phylogenetic regression, including ancestral character states (grey) reconstructed in the same trait space. **b**, Sperm length and SR length character state transitions from optimal trait-dependent Bayesian models. Arrow thickness and colour refer to transition probabilities with thicker lines and warmer colours corresponding to higher probability transitions. Adjacent numbers are model posterior probabilities resulting from rjMCMC. *Z* scores are the percentage of samples with a transition probability of 0. **c**, Phylogeny of *Drosophila* and kin with a branch colour gradient corresponding to the posterior probability density of 1,000 stochastic character mappings of female sperm-storage modes. The legend shows the probability of sperm being stored in both the SP and the SR ('SP + SR'; blue; determined to be the ancestral state) or the state of females storing sperm in the 'SR only' (red). **d**, Estimated evolutionary rates of sperm length derived from multirate BM models applied to stochastic trait mappings of female sperm-storage mode ($n$ = 1,000 simulations). Horizontal bars indicate median values, box bounds represent the first and third quartiles, and whiskers indicate 1.5× the interquartile range.

associated with sperm length were mutually exclusive from, and not in linkage disequilibrium with, those associated with SR length (average correlation coefficient ($R$) = 0.007; maximum $R$ = 0.04). Note, however, that preference–ornament linkage disequilibrium is not a prerequisite for good genes processes, and its absence in the DGRP lines does not preclude a historical role of runaway selection on sperm and SR length in *Drosophila*[23]. Some theoretical models further predict a bias toward sex-linkage of sex-specific traits, including sexually selected ornaments and preferences[26,27]. However, we found no significant X chromosome enrichment of genes for either trait ($\chi^2$ = 0.79, $P$ = 0.37 for sperm length; $\chi^2$ = 0.45, $P$ = 0.98 for SR length).

Our GWAS results were particularly amenable to testing the extensive and influential 'good genes' models, which collectively predict that ornament expression should be highly polygenic and not limited to genes for development of the ornament per se (that is, testis-specific or testis-enriched). Rather, ornament size reflects the bearer's (potentially genome-wide) 'condition', and thus indirect genetic benefits of mate choice (or sperm use[14]), because ornament expression is connected to organism-wide functionality across diverse molecular processes[22,28,29]. To test these a priori predictions with regard to sperm length evolution in *Drosophila*, we first determined the tissue of maximum expression for each ornament gene (Supplementary Data 4). Only 27 of the 142 genes (19.0%) exhibited maximum expression in the testis (Fig. 3a), and their average effect size was not significantly different from the remaining genes pleiotropically affecting sperm length (that is, maximum expression in tissues other than testis; $F_{5,137}$ = 1.02, $P$ = 0.40) (Supplementary Fig. 5). By contrast, other ornament genes exhibited maximum expression across a diversity of larval and adult tissues (Fig. 3a), including adult CNS (26.7%, which was a significantly greater contribution than testis; $\chi^2$ = 7.59, $P$ = 0.005), larval CNS (9.2%), eye (8.4%) and ovary (6.3%). The number of genes with enriched expression was significantly greater than expected for adult CNS (Fisher's exact test $P = 2.0 \times 10^{-3}$) (Fig. 3b) and eye (Fisher's exact test, $P$ = 0.03) (Fig. 3b). This pattern of outsized pleiotropic contribution to sperm length variation supports the 'genic capture' hypothesis[22,28,29].

Gene Ontology analyses of CNS sperm length genes revealed significant Biological Process enrichments, including nervous system development ($P = 2.5 \times 10^{-5}$), neuron differentiation ($P = 1.9 \times 10^{-3}$), axon development ($P = 1.04 \times 10^{-3}$), taxis ($P = 1.4 \times 10^{-4}$), synapse organization ($P = 3.6 \times 10^{-8}$), axonogenesis ($P = 9.8 \times 10^{-4}$), locomotion ($P = 5.4 \times 10^{-3}$) and chemotaxis ($P = 5.7 \times 10^{-4}$) (Supplementary Data 5). There was also a significant enrichment of immunoglobulin-like domain containing genes ($P = 1.5 \times 10^{-5}$), including a functionally integrated set that supports CNS connectivity. For example, three unlinked members of the

https://doi.org/10.1038/s41559-024-02587-2

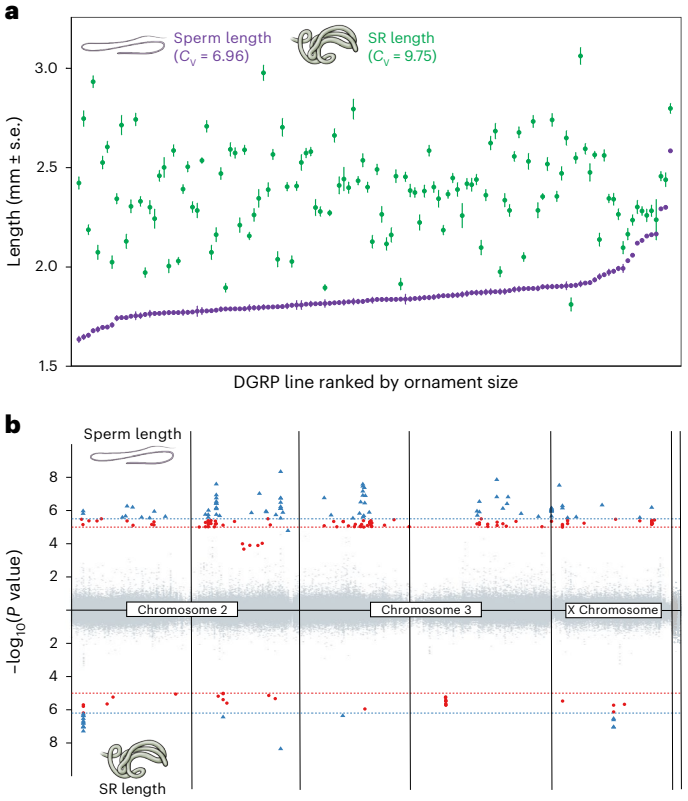


**Fig. 2 | Intraspecific phenotypic and genetic variation in sperm length and SR length are not correlated. a**, Mean ± s.e. values for sperm length (purple) and SR length (green) of 126 DGRP lines. The two traits are not significantly correlated among lines. **b**, Manhattan plot of significant variants associated with sperm length (upper) and SR length (lower). Genomic coordinates of variants and their associated significance levels are displayed. Significance is indicated for $P = 10^{-5}$ (red) and a trait-specific Benjamini–Hochberg false discovery rate threshold of 0.05 (blue).

*Dpr-interacting protein* gene family (*Dip*) were identified, which act as neuronal surface markers to mediate synaptic specificity through their interactions with *defective proboscis extension response* (*Dpr*) family members[30]. Multiple unlinked members of the *beaten* gene family and its interacting partner *sidestep* (*side*), which function in motor axon guidance, were also identified[31]. Finally, *Tenascin* accessory (*Ten*-a), which mediates axon growth and synaptic connectivity in the olfactory and neuromuscular systems, was also a candidate sperm length gene[32]. To further refine our understanding of CNS sperm length gene function, we surveyed expression across neuronal cell types. In total, 55.2% of these genes (16 of 29) exhibited maximum expression in cell types related to vision, including photoreceptor neurons, lamina monopolar neurons and amacrine neurons (Supplementary Fig. 6). As expected, *Dip-beta*, *Dip-delta* and *Ten-a* were among these genes.

Regarding testis-expressed sperm length genes, 37.0% (10 of 27) were long non-coding RNAs (lncRNAs). This proportion was greater than is present in the remaining 115 genes (13.4%; $\chi^2 = 15.27$, $P = 9.3 \times 10^{-5}$) and is consistent with the established role for lncRNAs in spermatogenesis in *Drosophila* and other organisms[33]. We utilized testis single-cell expression data to test the prediction that these sperm length genes exhibit post-meiotic expression associated with flagellum (that is, sperm length) development. As predicted, protein-coding genes increased markedly in expression in primary spermatocytes and remained upregulated through later stages of spermatid elongation (Fig. 3c). Although expressed at consistently lower levels than protein-coding genes, lncRNAs exhibited a strikingly similar pattern.

Theory regarding intersexual choice has focused on the benefits of mate discrimination for the female (either direct benefits, for example, resources, or indirect genetic benefits). Consequently, female preference evolution models generate predictions about genes underlying the preferred male traits, but not about the genetics of the female preference per se. With regard to premating discrimination, genes associated with the female sensory biology and cognition would be expected. By contrast, with regard to post-copulatory female choice and SR length, we predicted genes underlying the development and function of the female lower reproductive tract. This prediction was not met. In total, 19 genes were associated with SR length variants (Supplementary Data 4), including *rotund*, which encodes a transcription factor responsible for olfactory neuron specification[34], *short gastrulation*, a regulator of neuronal development through inhibition of bone morphogenic proteins[35], *terribly reduced optic lobes*, a regulator of cell signalling during development patterning decisions[36] and *Fish-lips*, a regulator of axon and dendrite targeting in olfactory receptor neurons[37]. Maximum expression analysis of a more inclusive set of SR length genes ($n = 146$; $P = 1.0 \times 10^{-4}$; Methods) revealed a significant enrichment of expression in adult CNS (observed = 40, expected = 21; $P = 3.0 \times 10^{-4}$) (Supplementary Fig. 7).

## Multiple conserved signalling pathways govern ornament elaboration

To further test theoretical predictions of indirect genetic benefit models for SR length evolution, we examined correlations between sperm length and other traits previously characterized in the identical DGRP lines. Most critically, theory predicts a positive relationship between male ornament size and female fitness[2,23]. This prediction was supported, because we found a significant positive correlation between sperm length and lifetime female fecundity[38] ($r = 0.21$, $P = 0.02$). With regard to nutrition and energy storage contributions to condition[22,28,29], we found significant positive associations between sperm length and both starvation resistance[25] ($r = 0.15$, $P = 0.045$) and attraction to ethyl acetate[39], which is a volatile compound produced by the fruit fly food source, yeast ($r = 0.23$, $P = 0.009$). By contrast, there were no significant correlations between sperm length and female lifespan ($r = 0.11$, $P = 0.23$), chill-coma recovery ($r = 0.074$, $P = 0.43$) or desiccation resistance ($r = -0.037$, $P = 0.69$). Finally, we identified a significant negative correlation between sperm length and male–male aggressive behaviour[40] ($r = -0.22$, $P = 0.015$), and note that sperm length gene *nervous wreck*, which regulates synaptic growth, was also identified in the study of aggressive behaviour[41]. Interpreting this relationship requires caution, however, because relationships between aggression and sex-specific fitness in *Drosophila* are complex[42].

In light of these phenotypic relationships, we next explored the insulin/IGF (IIS) pathway, which underlies horn development in dung beetles and has been postulated to be a common mechanism linking organism condition to the development of ornaments across diverse taxa, from insects to mammals[13,29]. We surveyed sex-specific targets of the *Drosophila* IIS pathway[40] and observed a significant enrichment of ornament genes among those that were upregulated in males (but not females) in an insulin receptor knockout background ($\chi^2 = 5.02$, $P = 0.02$). This included *Akt*, which encodes the core IIS kinase and is an essential regulator of tissue growth and body size[43]. The upstream activators of *Akt*, phosphatidylinositol phosphates, regulate cyst polarization and axoneme elongation during spermatogenesis[44]. To explore whether other central pathways governing development contribute to ornament elaboration[9,29], we determined whether sperm length genes were enriched among targets of the transcription factor *fruitless*, a regulator of neuronal sex differentiation[45] and in four conserved developmental pathways: *JAnus Kinase protein and the Signal Transducer and Activator of Transcription* (*JAK/STAT*), *Notch (N)*, *Decapentaplegic* (*dpp*) and *wnt/wingless*[46]. We found highly significant enrichments of our ornament genes in all of these pathways with the exception of

https://doi.org/10.1038/s41559-024-02587-2

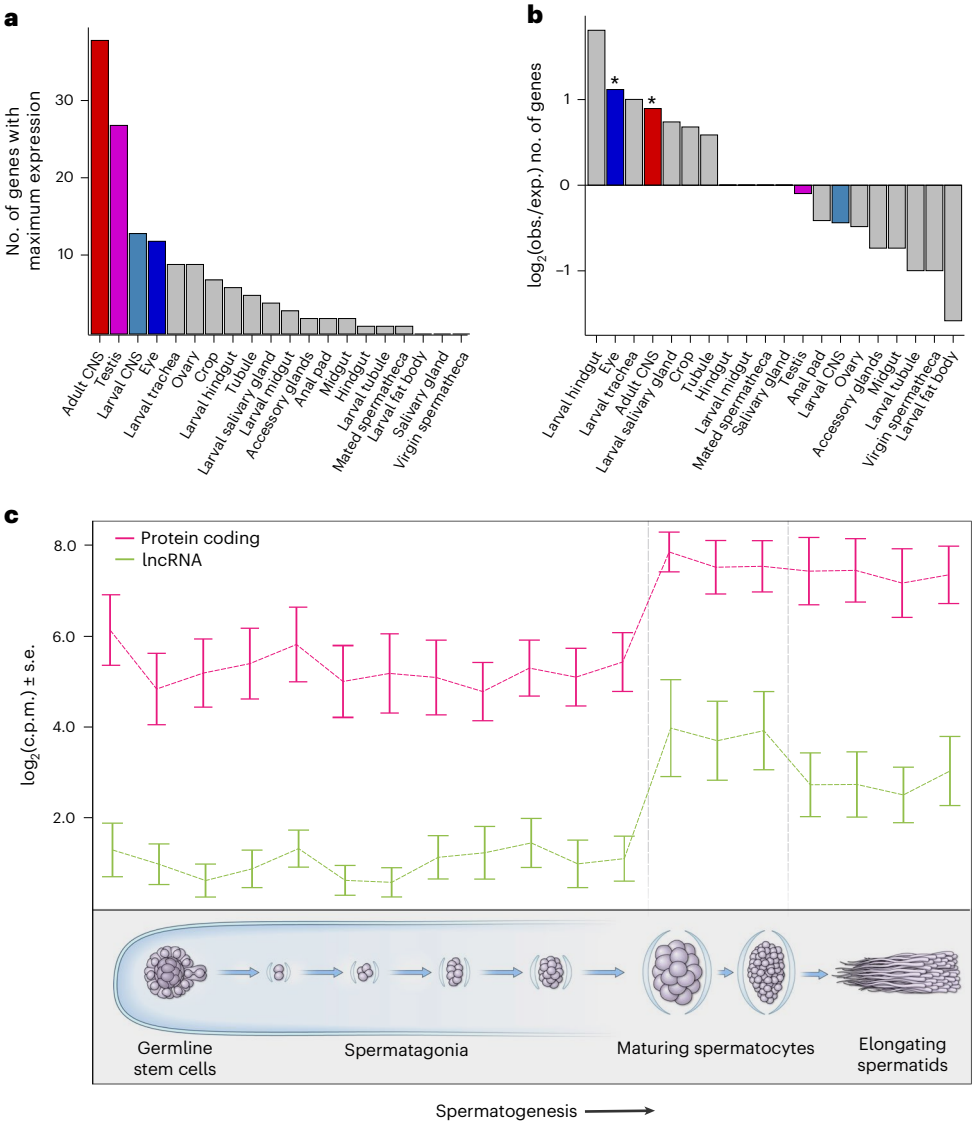


**Fig. 3 | Sperm length genes are biased towards expression in the CNS and sensory organs, as well as during the post-meiotic phase of spermatogenesis in the testes. a**, Distribution of sperm length candidate genes based on their maximum expression tissue. Adult CNS (red) has the highest expression followed by testis (purple), eye (dark blue) and larval CNS (light blue). **b**, Tissue-specific enrichment of genes (*y* axis) across adult and larval tissues (*x* axis). Two-sided Fisher's exact tests were used to assess significant enrichments ($P < 0.05$, asterisks). Adult CNS ($P = 2.0 \times 10^{-3}$; red) and eye ($P = 0.03$; dark blue) exhibited a significant enrichment above expectations. **c**, For genes with maximum expression in the testis, expression across spermatogenesis (*x* axis) is shown using mean (±s.e.) $\log_2$-transformed counts per million (c.p.m.; *y* axis). Both protein-coding genes ($n = 17$; magenta) and lncRNA genes ($n = 10$; green) showed an increase in expression levels at post-meiotic stages. lncRNA genes showed a decline in the latter stages of spermatogenesis. exp., expected; obs., observed.

*wnt/wingless* ($N$: $\chi^2 = 25.61$, $P = 5.0 \times 10^{-3}$; *fru*: $\chi^2 = 31.97$, $P = 4.0 \times 10^{-8}$; *JAK/STAT*: $\chi^2 = 13.31$, $P = 3.0 \times 10^{-4}$; *dpp*: $\chi^2 = 10.48$, $P = 1.0 \times 10^{-3}$). Sperm length genes also showed significant overrepresentation among multiple developmental pathways, including 82 genes in at least one pathway ($P = 0.01$), 37 genes in at least two pathways ($P = 1.0 \times 10^{-4}$), 13 genes in at least three pathways ($P < 1.0 \times 10^{-5}$) and of 5 genes in all four pathways ($P = 3.0 \times 10^{-5}$).

## Selective history of sperm length is consistent with the lek paradox

The 'lek paradox' is premised on the depletion of genetic variation for ornament elaboration in the face of intense and persistent selection mediated by female preference, which in turn would preclude females benefitting from being choosy. The 'genic capture' hypothesis provides a widely regarded resolution of the lek paradox[22]. It contends that the expression of ornaments will depend on the condition of their bearer in part because of the costs of growing and maintaining exaggerated traits[28,29]. Because condition itself will be determined by variation across many genes, a large mutational target space will maintain genetic variation in ornaments through mutation–selection balance[22]. The long sperm of *Drosophila* meet predictions of the genic capture hypothesis because they are costly to manufacture[15,17] and our GWAS results (above) revealed a genetic architecture characterized by many loci with diverse functions across the genome (Figs. 2b and 3a,c).

We further tested the applicability of sperm length evolution to the lek paradox by examining the effect of sperm length variants in relation to their recent selective history. The minor allele frequency (MAF) distribution of sperm length variants was significantly lower than that of SR length variants (Wilcoxon test, $P = 2.2 \times 10^{-16}$) and all 217 minor allele sperm length variants had a negative effect on sperm length, whereas only 57.7% (30 of 52) of SR length variants had a negative effect on SR length (Fig. 4a). Ancestral allele state reconstruction further revealed

https://doi.org/10.1038/s41559-024-02587-2

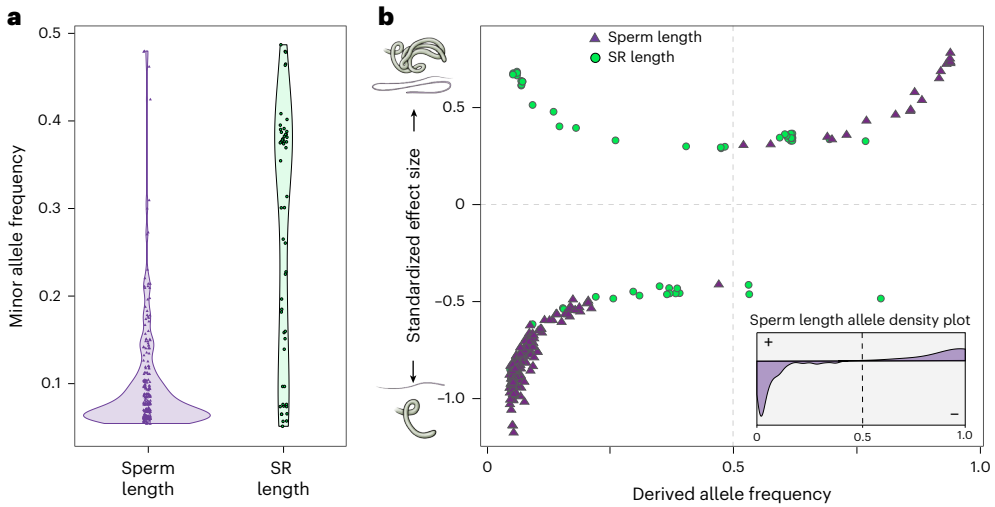


**Fig. 4 | Effect size distributions suggest different selective histories for genetic variants underlying sperm length and SR length expression.** **a**, Comparison of the MAF distributions for sperm length and SR length. **b**, Relationship between standardized effect size and derived allele frequency for variants associated with sperm length (purple triangles) and SR length (green circles). Inset, density plot of sperm length allele frequencies. It is noteworthy that high-frequency derived alleles have an exclusively positive effect on sperm length.

a striking pattern in which all high-frequency derived sperm length variants had positive effects on sperm length (Fig. 4b). This pattern was in stark contrast to the more balanced pattern observed for SR length variants (Fig. 4b) and supports the interpretation of recurrent positive selection on novel alleles that increase sperm length. Querying a haplotype screen for recent selective sweeps in the DGRP population[47], the majority of high-frequency derived sperm length alleles were in regions of sweeps (7 of 11; Fisher's exact test, $P = 0.05$), whereas low-frequency derived sperm length alleles did not exhibit an enrichment (41 of 90, $\chi^2 = 2.21$, $P = 0.13$) (Supplementary Fig. 8).

To further elucidate the adaptive significance of the negative effect sizes of sperm length variants, we compared their distribution with that of other traits that had previously been examined in the DGRP. The negative skew was shared by all stress resistance traits and by aggression, but not by other behavioural and life-history traits (Supplementary Fig. 9). This pattern is consistent with the interpretation that sperm length serves as a reliable, condition-dependent indicator to females of good genes that may contribute to offspring viability, with variation maintained through mutation–selection balance. Nevertheless, these comparisons should be interpreted with caution because they do not demonstrate a direct association.

### Relationship between intra- and interspecific genetic variation

We evaluated the genetics of sperm length and SR length diversification in a comparative genomic framework and compared results with the genetic architecture of *D. melanogaster*. Specifically, we examined evolutionary rate correlations between non-synonymous substitutions and integrated sperm and SR length (see Methods for justification) for 9,401 orthology groups across 15 species of *Drosophila* (Supplementary Data 6). This analysis identified 600 genes exhibiting significant correlated evolution, either positive or negative, with sperm and/or SR length (partial posterior probability <0.05 or >0.95). Of these genes, only 11 were also identified in our GWAS (Supplementary Table 1), which is consistent with random expectations (Fisher's exact test, $P = 0.36$). We therefore conclude that genes harbouring genetic variation governing sperm and/or SR length in *D. melanogaster* are largely distinct from those consistently associated with diversification in these traits across the phylogeny, which is consistent with previous studies[48,49]. However, it is noteworthy that 8 of the 11 overlapping genes (Supplementary Table 1) function in central developmental pathways, including the essential developmental signalling regulator *Notch* and an additional two genes that encode predominant components of the sperm flagellar axoneme (Supplementary Table 1).

Despite few overlapping genes, characteristics of gene function and expression were consistent between intra- and interspecific analyses. The latter revealed a strongly biased distribution with an order of magnitude more genes exhibiting significant negative than positive correlations with sperm and/or SR length (544 and 56 genes, respectively; binomial distribution, $P < 0.0001$). An examination of sex-biased expression revealed a statistically robust, opposing relationship. Specifically, a small set of male-biased genes exhibited accelerated evolution across lineages with exaggerated sperm and/or SR length, whereas a much larger set of female-biased genes exhibited accelerated evolution across lineages with reduced sperm and/or SR length (Fig. 5a). Positively correlated genes were skewed toward maximum expression in the testes (12 of 56; Fisher's exact test, $P = 0.001$) and genes encoding integral sperm protein components (13 of 56; Fisher's exact test, $P = 0.008$) (Fig. 5b), including a robust pattern of post-meiotic gene expression that closely parallels that of sperm length genes identified in the GWAS (Fig. 5c). Thus, persistent post-copulatory sexual selection drives the evolution of a small set of genes that directly influence sperm length (that is, non-pleiotropic, spermatogenesis genes) in a lineage-wide manner. By contrast, the much larger set of negatively correlated genes exhibited enriched ovary (263 of 544; 48.3%) (Fig. 5b) and larval CNS expression (254 of 544; 46.7%). The ovary-enriched genes included a significant number that are maternally contributed oocyte transcripts (165 of 263; $P = 1.6 \times 10^{-14}$) and are overwhelmingly detected during embryonic development (particularly in early stages 1–3). These genes comprised a significantly interconnected protein network (edge enrichment = 1.91-fold; $P = 8.4 \times 10^{-8}$) (Fig. 5d) characterized by functional enrichments relating to gene regulation, developmental lethality and abnormal neuroanatomy (Supplementary Data 6). As such, this core set of female-biased genes experiences intense purifying selection on high post-copulatory sexual selection lineages and accelerated evolution, perhaps because of relaxed constraints, when sperm length investment is reduced.

## Conclusions

Morphological and molecular interactions between sperm and the female reproductive tract are not well understood, but are recognized to be protracted and complex, and to generate selection on sperm

https://doi.org/10.1038/s41559-024-02587-2

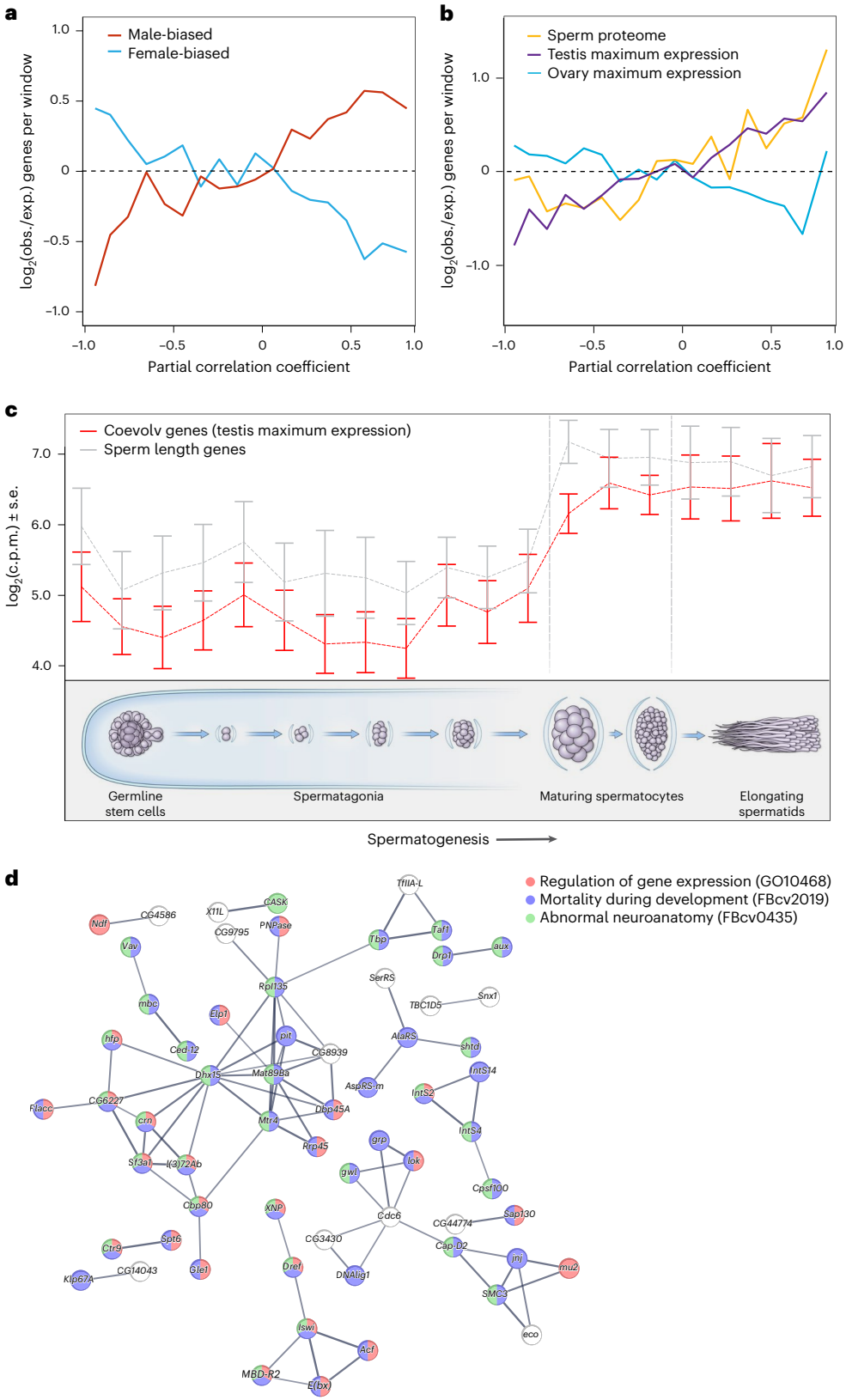


**Fig. 5 | Correlated gene evolution with ornament and/or preference diversification reveals functionally relevant expression patterns.** **a**, Male-biased (red) and female-biased (light blue) patterns of expression have opposite relationships with sperm and/or SR length evolution. **b**, Genes evolving in a correlated pattern with sperm and/or SR length were enriched for testis expression (purple) and encode integral sperm proteins (yellow). An opposing pattern was observed for ovary-enriched genes (light blue). **c**, For testis-enriched genes that evolve in a correlated manner with sperm length ($n = 12$; red), expression across spermatogenesis ($x$ axis) is displayed using mean (±s.e.) $\log_2$-transformed counts per million ($y$ axis). These genes exhibit patterns of post-meiotic spermatogenic expression that parallels that of the GWAS sperm length genes (grey; also see Fig. 3c). **d**, Network analysis of maternally contributed genes based on high confidence (>0.7) physical and functional interactions.

quantity and quality[20,50–53]. Post-copulatory sexual selection is credited with numerous aspects of sperm form exhibiting dramatic evolutionary diversification[7,16,54] and a taxonomically widespread pattern of co-diversification with critical dimensions of the female reproductive tract[50,55] The current, comparative analysis of 149 species reveals that sperm–SR length co-diversification in *Drosophila* is attributable to a process of co-evolution (Fig. 1). The striking rate disparity in sperm length evolution between female sperm-storage modalities (18 times higher in lineages storing sperm in the SR only versus those using both the SP and SR) was unexpected and highlights how incomplete our knowledge is of the proximate mechanisms of sperm–female interaction, even in this well-studied system[19,20]. Our understanding of ultimate selection pressures underlying post-copulatory (cryptic) female preferences (for example, how female fitness is enhanced through biasing fertilization in favour of certain sperm phenotypes) in any taxon, is even more rudimentary[4,5,56].

The evolution of a polygynandrous mating system in which males produce few giant sperm is paradoxical for evolutionary theory because of intrinsic negative self-reinforcement[15,57]. Because sperm length is negatively correlated with the number of sperm produced and transferred to females, as sperm evolve to be longer, there will be fewer sperm competing for each egg in the population[15,17,57]. This in turn should reduce the intensity of sexual selection for longer, more competitive sperm[57]. Results of the current study help to resolve this paradox by unveiling putative indirect genetic benefits to females of biasing fertilization in favour of longer sperm. Genetic variation in sperm length was positively correlated with lifetime female fecundity, starvation resistance and the ability to find food. We further revealed that genetic variation in sperm length contributes to a diverse repertoire of pleiotropic functions, including the neurogenomic basis of courtship behaviour and multimodal sensory processing, which could drive associations between success in premating and post-copulatory reproductive episodes.

The most rigorous and comprehensive theoretical treatment of 'condition' defines it as 'the relative capacity to maintain optimal functionality of essential cellular processes', and defines a 'condition-dependent display trait' as 'a conspicuous feature of an organism that varies in expression depending on the capacity to withstand environmental challenges'[28]. The demonstrated involvement of several central hormonal and developmental pathways (IIS, *N*, *Fru*, *dpp* and *JAK/STAT*) in sperm growth may provide the links between ornament production (growth of the sperm flagellum) and functionality of whole-organism systems[13,28,29]. Moreover, because development of the machinery of sperm production (testes) transcends both larval and adult life-history stages, sperm length may provide a cumulative signal to females of male genome quality and organismal vitality across environmentally disparate life stages. It is important to note, however, that pleiotropic relationships are complex, and knowledge of the relationship between sperm length and net lifetime fitness is beyond the scope of this investigation[58].

Consistent with theory[22], our data support a history of cyclic sexual selection in which new sperm length-enhancing variants selectively sweep through the population via indirect selection on SR length. The negatively biased effect size distribution for sperm length variants highlights the widespread pleiotropic relationships in which many genes can simultaneously have deleterious effects on ornament expression and organismal condition. These observations may also explain why so few of the sperm length genes identified in the intraspecific (GWAS) analysis in *D. melanogaster* were identified as contributing to sperm length diversification across the clade. Because of ongoing selection, both positive and negative, intraspecific analyses such as ours will capture only a restricted set of sperm length-enhancing variants on their way to fixation at a given time and a much larger set of deleterious alleles that are destined to be purged from the population[12].

A striking observation from both intraspecific and interspecific analyses of sperm and SR length genetics was the relationship between these sexually selected traits and nervous system development and function. The involvement of CNS genes is intuitive for premating but not post-copulatory female preferences, and perhaps for male behavioural (for example, courtship) but not structural ornaments. However, neural development and processing is probably one of the most demanding cellular processes[59], and suboptimal nervous system and cognitive performance is likely to erode fitness. We thus postulate taxonomically widespread selection for female preferences that discriminate among males based on any traits that reliably indicate genetic variation in nervous system condition[28]. Although the underlying reasons are not well understood, brains and testes (and neurons and sperm) exhibit similar gene expression and proteomic profiles[60]. Sperm quality may be a robust indicator of brain and nervous system quality, thereby explaining the dramatic evolution of sperm form throughout the animal kingdom[16,54].

## Methods

### Phylogenetic comparative analyses

Mean sex-specific body size (thorax length), sperm length and SR length ($n = 5$ per trait) were determined for 149 species of Drosophilidae from four genera: *Drosophila*, *Scaptodrosophila*, *Chymomyza* and *Zaprionus* (Supplementary Data 1) using methods reported in ref. 61. Cultures were acquired from the National *Drosophila* Species Stock Center at Cornell University (stock numbers are provided in Supplementary Data 1) or from the Artyom Kopp laboratory at the University of California at Davis. Culturing methods are provided in ref. 61. Sperm length data for 37 species were previously reported in refs. 61,62. SR data for 28 species were previously reported in ref. 61. Sperm length and SR length data are reported here for 117 and 121 species, respectively. Sperm-storage organ usage ('SR only' or 'SP + SR'; 'sperm-storage mode') was determined by visualizing sperm in inseminated females using differential interference contrast microscopy.

Phylogenetic comparative analyses of log-transformed continuous data on sperm length, SR length and body size were conducted using a multilocus phylogenetic tree of 360 drosophilid species[63] pruned to include only the 149 examined species. Ancestral sperm length and SR length were reconstructed using the functions 'phenogram', 'contMap' and 'phylomorphospace' in phytools v.1.9-16 (ref. 64). We compared a set of PGLS models using the function 'pgls' in caper v.1.0.2 while estimating $\lambda$ using maximum likelihood[65]. Male and female body size and sperm-storage organ use were included as covariates in PGLS models, and model comparison clearly determined that including adult male thorax length as a covariate is optimal (difference in Akaike information criterion ($\Delta$AIC) > 2 for the next best model).

We compared evolutionary models of trait dependency between (1) sperm length and SR length and (2) sperm-storage mode and SR length using reversible-jump Markov chain Monte Carlo (rjMCMC) and the 'discrete' program of BayesTraits v.2 (refs. 66,67) after rendering the continuous ornament and preference trait data discrete by splitting their log-transformed distributions at the median. Independent and dependent models were compared with results respectively combined from three independent rjMCMC runs yielding a total of 3,000 samples of the posterior (1,000 per run) initiated from different starting seeds to ensure proper sampling of the posterior distribution. Each chain ran for 21,000,000 iterations, sampling every 20,000 generations after discarding the first 1,000,000 iterations of each chain as burn-in. Model comparison indicated that dependent models fit the data substantially better than independent models (that is, a model in which the probability of character state transitions in sperm length and SR length are independent of each other).

Ancestral female sperm-storage modes ('SP + SR' versus 'SR only') were reconstructed using stochastic character mapping with phytools[64] and 1,000 simulations under transition models allowing

different rates. The impact of sperm-storage mode on the rate of sperm length evolution was examined by comparing the fit of one-rate and two-rate Brownian motion (BM) models (simple macro-evolutionary models with either shared or separate rates per mode) with a variety of more complex Ornstein–Uhlenbeck models (that is, models featuring trait optima and stationary selection) using OUwie v.2.10 (ref. 68). Models were compared across the set of 1,000 stochastic character mappings of sperm-storage mode and included our species error data for sperm length, and we found clear support for multirate BM models in our observed data (ΔAIC = 2.9–3.2 for single-rate BM models across 99% of stochastic mappings of mode). To further validate our model fitting results, we simulated data using 'OUwie.sim' with the mode-specific estimates of $\sigma^2$ (instantaneous sigma) from two-rate BM models. Model comparison showed that we can unambiguously distinguish between a single-rate and multirate BM model fitted to simulated data (ΔAIC > 70 for single-rate BM models).

### GWAS of sperm and SR length in *D. melanogaster*

Identical methods to those described above were used to quantify intraspecific variation in body size, sperm length and SR length in 126 lines of the DGRP[25] obtained from the Bloomington *Drosophila* Stock Center. To minimize environmental effects, flies from each line were reared by transferring 70 first-instar larvae to polystyrene vials containing 8 ml of National *Drosophila* Species Stock Center cornmeal culture medium and maintaining them at 22.5 °C in a photoperiod-controlled growth chamber. Adult males and females were collected on the day of eclosion and kept in sex-specific groups of 10 flies in vials with media and live yeast for 5 days before dissection for measurement of sperm or SR length. Mean sperm length for each male was calculated across the five longest of seven sperm measured (to minimize underestimates attributable to sperm breakage). Mean sperm and SR lengths were based on 12 males and 16 females per line, respectively. The genome-wide association (GWA) analyses (described below) are thus based on length measurements of 7,560 sperm from 1,512 males (five sperm per male) and SRs from 2,016 females.

We estimated among-line variance ($\sigma_L^2$) and environmental variance ($\sigma_e^2$) components, broad-sense heritability ($H^2$) and the interline $C_V$ for each trait as in ref. 69. Based on a Tukey's outlier test, we identified as outliers four lines (~3% of the data) with the longest sperm lengths. We Winsorized the mean sperm length of these lines to the mean sperm length of the fifth longest line before using them for GWA analyses to account for possible spurious outlier effect while minimizing the loss of biological information in our dataset. With regard to SR length, the distribution of line means did not deviate from normality and there were no significant outliers. We performed separate GWA analyses (dgrp2.gnets.ncsu.edu/) for sperm length and SR length to estimate the additive effects of 1,884,703 common variants (MAF) ≥ 0.05)[70] across 126 DGRP lines using a mixed linear model estimate that accounts for between-line relatedness after adjusting for inversions and *Wolbachia* infection[70]. A significance threshold of $P < 10^{-5}$ was applied given its common usage in previous DGRP studies (for example, refs. 41,70). We also report results after Benjamini–Hochberg multiple test corrections. A quantile–quantile plot of *P* values (Supplementary Fig. 10) revealed a disparity in statistical power between the traits. To explore the contribution of effect size to this disparity, we estimated the phenotypic variance explained by variants using a genomic best linear unbiased prediction model[71] implemented in the R package qgg[72]. This revealed that significant variants ($P \leq 10^{-5}$) explained 69% of sperm length variation but only 31% of SR length. Relaxation of the *P* value threshold ($P \leq 10^{-4}$) raised the explanatory power for SR length by 18% and included 204 additional variants. Given the comparable trait heritability, we conclude that SR length is determined by more variants of modest effect size. In addition, measurement precision can affect GWAS power[73]. We therefore estimated error variance (that is, the variance not attributable to line or individual) using the VCA[74]. This revealed an estimated 28.7% higher error for SR length (estimate 5.58, 95% confidence interval 5.42–5.77) than for sperm length (estimate 3.8, 95% confidence interval 3.91–4.04)

The MAF and effect size ('SingleEff') were obtained from the GWA linear mixed model (Supplementary Data 2). Ancestral allele state was determined for single nucleotide polymorphisms using maximum likelihood (Kimura substitution model) as implemented by the est-sfs program[75]. This analysis was based upon the *D. melanogaster* (BDGP R5/dm3), *Drosophila simulans*, *Drosophila yakuba and Drosophila eugracilis* genomes with *Drosophila subpulchrella* as an outgroup to polarize alignments. A parsimony-based approach was used for indel variation based upon the *D. melanogaster* (BDGP R5/dm3), *D. simulans*, *D. yakuba and Drosophila erecta* genomes.

Linkage disequilibrium was estimated between significant sperm and SR variants using PLINK[76]; the highest linkage disequilibrium between any two variants was 0.04. The DGRP genome assembly (BDGP R5/dm3) was used to identify candidate genes neighbouring significant variants associated with sperm and SR length. For intergenic variants, genes in a 10 kb window in either direction were considered candidate genes. Where no genes were within 10 kb, the nearest single gene was considered a candidate. Differences in allele frequency and effect size distributions between sperm length and SR length were examined using Wilcoxon rank-sum tests.

Correlations between sperm length and previously characterized phenotypes in the identical DGRP lines were conducted after exhaustively mining the literature for studies with a minimum of 100 DGRP lines in common with the current study. Relevant phenotypes included female lifespan and lifetime fecundity[38] (118 lines in common), male aggressive behaviour[41] (126 lines in common), chill-coma recovery, desiccation and starvation resistance[25] (118 lines in common), and attraction to ethanol and ethyl acetate[39] (126 lines in common). Pearson correlation coefficients between the Winsorized sperm length and the phenotypes mentioned above were calculated using R function 'cor' following Benjamini–Hochberg adjustment for multiple testing using R function 'p.adjust'[77]. We assessed the effect size distributions of other traits (stress resistance, behavioural, life history) previously examined using the DGRP to identify traits with negative effect size distributions comparable with sperm length (see Supplementary Fig. 9 for traits and citations).

### Gene expression analyses

*Drosophila melanogaster* tissue-specific RNA sequencing data were obtained from FlyAtlas 2 (ref. 78). Maximum expression of a gene was determined after *Z*-score normalization of transcripts per million using the 'scale' function[77]. Adult brain and thoracico-abdominal ganglion were averaged to create a single 'Adult CNS' category. Over- and under-representation of maximum expression in each tissue were tested using $\chi^2$ (with Yates' correction) using a null expectation established from the genome-wide pattern of maximum expression for each tissue. Highly comparable results were obtained using patterns of tissue enriched expression (more than twofold expression whole body). Gene Ontology and gene enrichment analyses were performed using FlyMine[79]. The effect size of minor alleles associated with genes in different tissue expression categories was compared using a linear model. *Drosophila melanogaster* single-cell expression data for testis and CNS were obtained from the annotated dataset available on Fly Cell Atlas[80]. For CNS, average expression was used for each annotated cluster to determine maximum expression. For testis, average expression for non-somatic clusters was ordered temporally according to spermatogenic stages.

### Enrichment of ornament genes in conserved developmental pathways

We tested whether sperm length genes were enriched in the *IIS* and five additional conserved developmental pathways: *fruitless*, *Notch*, *JAK/STAT*, *Decapentaplegic* and *wnt/wingless*. Comparisons were restricted

to studies that surveyed 80% or more of *D. melanogaster* genes[40,45,81–84]. For IIS, comparisons were conducted separately for differentially expressed genes in a background expressing a dominant-negative insulin-like receptor for males (348 upregulated and 801 downregulated genes) and females (26 upregulated and 19 downregulated genes) brain[40]. *Fruitless* is involved in *Drosophila* sex determination, is expressed in a highly specialized set of neurons[85] and male-specific splicing variant $Fru^M$ is required for male courtship behaviour[86,87]. *Fru* targets were identified by overexpressing $Fru^M$ in *Fru*-expressing neurons[45]. *Notch* signalling is essential for cell fate during embryonic development and the maintenance of progenitor stem cells[88]. *Notch* pathway members were based upon the RNAi screen of Mummery-Widmer et al.[81]. *JAK/STAT* signalling is involved in morphogenetic processes during embryonic development and interacts with the IIS through competitive inhibition[89–91]. Canonical and non-canonical targets of *JAK/STAT* from *JAK* gain- and loss-of-function and *STAT* loss-of-function mutants during embryogenesis (twofold or greater change in expression) were used[82]. *Decapentaplegic* (*Dpp*), a transforming growth factor β homologue in *Drosophila*, signalling pathway functions in several stem cell niches including the germline. *Dpp* target genes are indirectly induced by suppressing *brinker* (*brk*), which is a repressor of *Dpp* pathways genes[92], chromatin immunoprecipitation followed by sequencing binding sites were used as the basis for the identification of *Dpp* target genes[83].

*Wnt/wingless* (*Wnt/wg*) signalling contributes to development (for example, axon guidance, body segmentation, limb development and stem cell differentiation) and adult metabolism (for example, adipogenesis and triglyceride metabolism)[84,93,94]. A genome-wide RNA interference screen in imaginal discs with subsequent validation was used for comparison[84]. Gene enrichment across these pathways and exact significance levels were determined by random sampling (100,000 iterations).

### Detection of selective sweeps

Surveys for selective sweeps in the DGRP population were conducted using the H12 model data from ref. 47. A conservative expected distribution of H12 values under constant $N_e = 10^6$ and a recombination rate of $5 \times 10^{-7}$ cM per bp were simulated to establish a one-per-genome false detection rate level. Scanning windows of 400 single nucleotide polymorphisms containing H12 values above the cut-off were used to assign evidence of a selective sweep. Significant variants were mapped onto the H12 scan windows. If variants belonged to multiple windows, the windows were merged and the highest H12 value was retained. Fisher's exact tests were used to establish statistical significance relative to the background prevalence of sweeps across the genome separately for high- and low-frequency derived alleles.

### Comparative genomic analysis of ornament size/preference strength

We examined evolutionary rate correlations between non-synonymous substitutions and the combined ornament size and/or preference strength expression using the program Coevol[95]. Although we exclusively used data for sperm length, the robust pattern of co-diversification in ornament size and preference strength (Fig. 1a) means that almost identical results would derive from using preference strength data. We thus interpret the results as genetic correlates of the combined ornament and/or preference character states. Annotated protein sequences for six *Drosophila melanogaster* subgroup species (*D. melanogaster*, *D. simulans*, *Drosophila sechellia*, *D. erecta*, *D. yakuba* and *Drosophila ananassae*) were obtained from FlyBase and supplemented with an additional nine species from the National Center for Biotechnology Information: *Drosophila biarmipes* (GCF_000233415.1), *Drosophila bipectinata* (GCF_000236285.1), *Drosophila elegans* (GCF_000224195.1), *D. eugracilis* (GCF_000236325.1), *Drosophila ficusphila* (GCF_000220665.1), *Drosophila kikkawai* (GCF_000224215.1), *Drosophila rhopaloa* (GCF_000236305.1), *Drosophila suzukii* (GCF_000472105.1) and *Drosophila takahashii* (GCF_000224235.1). Orthology was determined using a local installation of OrthoDB[96]. The resulting orthology relationships formed the basis for producing protein-based alignments using the MAFFT L-INS-I algorithm, which were subsequently back-translated[97]. The Coevol analysis was restricted to those orthology groups with 1:1 orthology relationships in at least 14 of the 15 species. Ancestral sperm length was constructed using the FastAnc function from the phytools package[64]. The Coevol (1.6) program[95] was run in the dN mode using two runs per orthology group and convergence was assessed using the tracecomp program (minimum effective size >300 and a maximum difference <0.1). The converged runs were analysed using readcoevol to obtain partial correlation coefficients and partial posterior probabilities between dN and sperm length after controlling for dS. Syracuse University's OrangeGrid HTCondor distributed computing platform was used to run 6,000 concurrent Coevol analyses in intervals of 10,000 steps requiring over 15 million computing hours. After each interval, convergence was assessed using the tracecomp program. In total, convergence was achieved for 9,401 orthology groups. Network interaction analyses were determined using high confidence (>0.7) physical and functional interactions in the STRING database[98].

Unless otherwise mentioned, all phylogenetic and statistical analyses were performed using R statistical software[77]. All statistical tests of null hypotheses were two-tailed.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All data are available in the Supplementary Information accompanying this paper.

## References

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* Vol. 2 (John Murray, 1871).
2. Andersson, M. B. *Sexual Selection* (Princeton Univ. Press, 1994).
3. Parker, G. A. Sperm competition and its evolutionary consequences in the insects. *Biol. Rev.* **45**, 525–567 (1970).
4. Eberhard, W. G. *Female Control: Sexual Selection by Cryptic Female Choice* (Princeton Univ. Press, 1996).
5. Keller, L. & Reeve, H. K. in *Advances in the Study of Behavior* Vol. 24 (eds Slater, P. J. B. et al.) 291–315 (Elsevier, 1995).
6. Birkhead, T., Hosken, D. J. & Pitnick, S. (eds) *Sperm Biology: An Evolutionary Perspective* (Elsevier, 2009).
7. Lüpold, S. & Pitnick, S. Sperm form and function: what do we know about the role of sexual selection? *Reproduction* **155**, R229–R243 (2018).
8. Chenoweth, S. F. & McGuigan, K. The genetic basis of sexually selected variation. *Annu. Rev. Ecol. Evol. Syst.* **41**, 81–101 (2010).
9. Wilkinson, G. S. et al. The locus of sexual selection: moving sexual selection studies into the post-genomics era. *J. Evol. Biol.* **28**, 739–755 (2015).
10. Johnston, S. E. et al. Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93–95 (2013).
11. Barmina, O. & Kopp, A. Sex-specific expression of a HOX gene associated with rapid morphological evolution. *Dev. Biol.* **311**, 277–286 (2007).
12. Parrett, J. M. et al. Genomic evidence that a sexually selected trait captures genome-wide variation and facilitates the purging of genetic load. *Nat. Ecol. Evol.* **6**, 1330–1342 (2022).
13. Emlen, D. J., Warren, I. A., Johns, A., Dworkin, I. & Lavine, L. C. A mechanism of extreme growth and reliable signaling in sexually selected ornaments and weapons. *Science* **337**, 860–864 (2012).

14. Yasui, Y. A 'good-sperm' model can explain the evolution of costly multiple mating by females. *Am. Nat.* **149**, 573–584 (1997).
15. Lüpold, S. et al. How sexual selection can drive the evolution of costly sperm ornamentation. *Nature* **533**, 535–538 (2016).
16. Kahrl, A. F., Snook, R. R. & Fitzpatrick, J. L. Fertilization mode drives sperm length evolution across the animal tree of life. *Nat. Ecol. Evol.* **5**, 1153–1164 (2021).
17. Pitnick, S. Investment in testes and the cost of making long sperm in *Drosophila*. *Am. Nat.* **148**, 57–80 (1996).
18. Miller, G. T. & Pitnick, S. Sperm–female coevolution in *Drosophila*. *Science* **298**, 1230–1233 (2002).
19. Manier, M. K. et al. Resolving mechanisms of competitive fertilization success in *Drosophila melanogaster*. *Science* **328**, 354–357 (2010).
20. Lüpold, S. et al. How female × male and male × male interactions influence competitive fertilization in *Drosophila melanogaster*. *Evol. Lett.* **4**, 416–429 (2020).
21. Miller, G. T. & Pitnick, S. Functional significance of seminal receptacle length in *Drosophila melanogaster*. *J. Evol. Biol.* **16**, 114–126 (2003).
22. Tomkins, J. L., Radwan, J., Kotiaho, J. S. & Tregenza, T. Genic capture and resolving the lek paradox. *Trends Ecol. Evol.* **19**, 323–328 (2004).
23. Kirkpatrick, M. & Ryan, M. J. The evolution of mating preferences and the paradox of the lek. *Nature* **350**, 33–38 (1991).
24. Pitnick, S., Marrow, T. & Spicer, G. S. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. *Evolution* **53**, 1804–1822 (1999).
25. Mackay, T. F. C. et al. The *Drosophila melanogaster* genetic reference panel. *Nature* **482**, 173–178 (2012).
26. Kirkpatrick, M. & Hall, D. W. Sexual selection and sex linkage. *Evolution* **58**, 683–691 (2004).
27. Connallon, T. & Clark, A. G. Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* **64**, 3417–3442 (2010).
28. Hill, G. E. Condition-dependent traits as signals of the functionality of vital cellular processes: condition is more than nutrition. *Ecol. Lett.* **14**, 625–634 (2011).
29. Warren, I. A., Gotoh, H., Dworkin, I. M., Emlen, D. J. & Lavine, L. C. A general mechanism for conditional expression of exaggerated sexually-selected traits. *Bioessays* **35**, 889–899 (2013).
30. Xu, S. et al. Interactions between the Ig-Superfamily proteins *DIP-α* and *Dpr6/10* regulate assembly of neural circuits. *Neuron* **100**, 1369–1384.e6 (2018).
31. Aberle, H. Searching for guidance cues: follow the *Sidestep* trail. *Fly (Austin)* **3**, 270–273 (2009).
32. Mosca, T. J. & Luo, L. Synaptic organization of the *Drosophila* antennal lobe and its regulation by the Teneurins. *eLife* **3**, e03726 (2014).
33. Wen, K. et al. Critical roles of long noncoding RNAs in *Drosophila* spermatogenesis. *Genome Res.* **26**, 1233–1244 (2016).
34. Li, Q., Barish, S., Okuwa, S. & Volkan, P. C. Examination of endogenous *rotund* expression and function in developing *Drosophila* olfactory system using CRISPR-Cas9–mediated protein tagging. *G3 (Bethesda)* **5**, 2809–2816 (2015).
35. Biehs, B., François, V. & Bier, E. The *Drosophila short gastrulation* gene prevents Dpp from autoactivating and suppressing neurogenesis in the neuroectoderm. *Genes Dev.* **10**, 2922–2934 (1996).
36. Guss, E. J., Akbergenova, Y., Cunningham, K. L. & Littleton, J. T. Loss of the extracellular matrix protein Perlecan disrupts axonal and synaptic stability during *Drosophila* development. *eLife* **12**, RP88273 (2023).
37. Xie, Q. et al. Transsynaptic *Fish-lips* signaling prevents misconnections between nonsynaptic partner olfactory neurons. *Proc. Natl Acad. Sci. USA* **116**, 16068–16073 (2019).
38. Durham, M. F., Magwire, M. M., Stone, E. A. & Leips, J. Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. *Nat. Commun.* **5**, 4338 (2014).
39. Arya, G. H. et al. The genetic basis for variation in olfactory behavior in *Drosophila melanogaster*. *Chem. Senses* **40**, 233–243 (2015).
40. Graze, R. M., Tzeng, R.-Y., Howard, T. S. & Arbeitman, M. N. Perturbation of IIS/TOR signaling alters the landscape of sex-differential gene expression in *Drosophila*. *BMC Genomics* **19**, 893 (2018).
41. Shorter, J. et al. Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior. *Proc. Natl Acad. Sci. USA* **112**, E3555–E3563 (2015).
42. Kravitz, E. A., Fernandez, M. & de la, P. Aggression in *Drosophila*. *Behav. Neurosci.* **129**, 549–563 (2015).
43. Oldham, S. & Hafen, E. *Insulin/IGF* and target of rapamycin signaling: a *TOR* de force in growth control. *Trends Cell Biol.* **13**, 79–85 (2003).
44. Wei, H.-C. et al. Depletion of plasma membrane *PtdIns(4,5) P2* reveals essential roles for phosphoinositides in flagellar biogenesis. *J. Cell Sci.* **121**, 1076–1084 (2008).
45. Dalton, J. E. et al. Male-specific *Fruitless* isoforms have different regulatory roles conferred by distinct zinc finger DNA binding domains. *BMC Genomics* **14**, 659 (2013).
46. Babonis, L. S. & Martindale, M. Q. Phylogenetic evidence for the modular evolution of metazoan signalling pathways. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20150477 (2017).
47. Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet.* **11**, e1005004 (2015).
48. Gleason, J. M. & Ritchie, M. G. Do quantitative trait loci (QTL) for a courtship song difference between *Drosophila simulans* and *D. sechellia* coincide with candidate genes and intraspecific QTL? *Genetics* **166**, 1303–1311 (2004).
49. Arbuthnott, D. The genetic architecture of insect courtship behavior and premating isolation. *Heredity* **103**, 15–22 (2009).
50. Pitnick, S., Wolfner, M. F. & Suarez, S. S. in *Sperm Biology: An Evolutionary Perspective* (eds Birkhead, T. R. et al.) 247–304 (Elsevier, 2009).
51. Immler, S. et al. Resolving variation in the reproductive tradeoff between sperm size and number. *Proc. Natl Acad. Sci. USA* **108**, 5325–5330 (2011).
52. Pitnick, S., Wolfner, M. F. & Dorus, S. Post-ejaculatory modifications to sperm (PEMS). *Biol. Rev.* **95**, 365–392 (2020).
53. McCullough, E. L. et al. The life history of *Drosophila* sperm involves molecular continuity between male and female reproductive tracts. *Proc. Natl Acad. Sci. USA* **119**, e2119899119 (2022).
54. Pitnick, S., Hosken, D. J. & Birkhead, T. R. in *Sperm Biology: An Evolutionary Perspective* (eds Birkhead, T. R. et al.) 69–149 (Elsevier, 2009).
55. Higginson, D. M., Miller, K. B., Segraves, K. A. & Pitnick, S. Female reproductive tract form drives the evolution of complex sperm morphology. *Proc. Natl Acad. Sci. USA* **109**, 4538–4543 (2012).
56. Firman, R. C., Gasparini, C., Manier, M. K. & Pizzari, T. Postmating female control: 20 years of cryptic female choice. *Trends Ecol. Evol.* **32**, 368–382 (2017).
57. Bjork, A. & Pitnick, S. Intensity of sexual selection along the anisogamy–isogamy continuum. *Nature* **441**, 742–745 (2006).
58. Radwan, J., Engqvist, L. & Reinhold, K. A paradox of genetic variance in epigamic traits: beyond "good genes" view of sexual selection. *Evol. Biol.* **43**, 267–275 (2016).
59. Götz, M. & Huttner, W. B. The cell biology of neurogenesis. *Nat. Rev. Mol. Cell Biol.* **6**, 777–788 (2005).

60. Matos, B., Publicover, S. J., Castro, L. F. C., Esteves, P. J. & Fardilha, M. Brain and testis: more alike than previously thought? *Open Biol.* **11**, 200322 (2021).
61. Syed, Z. A. et al. Sperm cyst 'looping': a developmental novelty enabling extreme male ornament evolution. *Cells* **10**, 2762 (2021).
62. Pitnick, S., Markow, T. A. & Spicer, G. S. Delayed male maturity is a cost of producing large sperm in *Drosophila*. *Proc. Natl Acad. Sci. USA* **92**, 10614–10618 (1995).
63. Kim, B. Y. et al. Single-fly genome assemblies fill major phylogenomic gaps across the *Drosophilidae* tree of life. *PLoS Biol.* **22**, e3002697 (2024).
64. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
65. Orme, D. et al. The caper package: comparative analysis of phylogenetics and evolution in R. R package version 5 (2013).
66. Pagel, M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. Biol. Sci.* **255**, 37–45 (1994).
67. Pagel, M. & Meade, A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167**, 808–825 (2006).
68. Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C. & O'Meara, B. C. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369–2383 (2012).
69. Falconer, D. S. & Mackay, T. *Introduction to Quantitative Genetics* (Pearson, 2009).
70. Mackay, T. F. C. & Huang, W. Charting the genotype–phenotype map: lessons from the *Drosophila melanogaster* genetic reference panel. *Wiley Interdiscip. Rev. Dev. Biol.* **7**, 10.1002/wdev.289 (2018).
71. Ober, U. et al. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1002685 (2012).
72. Rohde, P. D., Fourie Sørensen, I. & Sørensen, P. qgg: an R package for large-scale quantitative genetic analyses. *Bioinformatics* **36**, 2614–2615 (2020).
73. Liao, J. et al. Impact of measurement error on testing genetic association with quantitative traits. *PLoS ONE* **9**, e87044 (2014).
74. Schuetzenmeister, A. & Dufey, F. VCA: Variance Component Analysis. R package version 1.5.1 (2022).
75. Keightley, P. D. & Jackson, B. C. Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics* **209**, 897–906 (2018).
76. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
77. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
78. Leader, D. P., Krause, S. A., Pandit, A., Davies, S. A. & Dow, J. A. T. FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res.* **46**, D809–D815 (2018).
79. Lyne, R. et al. FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics. *Genome Biol.* **8**, R129 (2007).
80. Li, H. et al. Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022).
81. Mummery-Widmer, J. L. et al. Genome-wide analysis of *Notch* signalling in *Drosophila* by transgenic RNAi. *Nature* **458**, 987–992 (2009).
82. Tsurumi, A., Zhao, C. & Li, W. X. Canonical and non-canonical *JAK/STAT* transcriptional targets may be involved in distinct and overlapping cellular processes. *BMC Genomics* **18**, 718 (2017).
83. Doumpas, N. et al. *Brk* regulates wing disc growth in part via repression of *Myc* expression. *EMBO Rep.* **14**, 261–268 (2013).
84. DasGupta, R., Kaykas, A., Moon, R. T. & Perrimon, N. Functional genomic analysis of the *wnt-wingless* signaling pathway. *Science* **308**, 826–833 (2005).
85. Kimura, K.-I., Hachiya, T., Koganezawa, M., Tazawa, T. & Yamamoto, D. *Fruitless* and *doublesex* coordinate to generate male-specific neurons that can initiate courtship. *Neuron* **59**, 759–769 (2008).
86. Ryner, L. C. et al. Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* **87**, 1079–1089 (1996).
87. Wohl, M., Ishii, K. & Asahina, K. Layered roles of *fruitless* isoforms in specification and function of male aggression-promoting neurons in *Drosophila*. *eLife* **9**, e52702 (2020).
88. Zhou, B. et al. *Notch* signaling pathway: architecture, disease, and therapeutics. *Signal Transduct. Target. Ther.* **7**, 95 (2022).
89. Hombría, J. C.-G. & Sotillos, S. *JAK-STAT* pathway in *Drosophila* morphogenesis: from organ selector to cell behavior regulator. *JAKSTAT* **2**, e26089 (2013).
90. Rosendo Machado, S., van der Most, T. & Miesen, P. Genetic determinants of antiviral immunity in dipteran insects – compiling the experimental evidence. *Dev. Comp. Immunol.* **119**, 104010 (2021).
91. Meng, Q., Xu, Y., Li, Y. & Wang, Y. Novel studies on *Drosophila melanogaster* model reveal the roles of *JNK–Jak/STAT* axis and intestinal microbiota in insulin resistance. *J. Drug Target.* **31**, 261–268 (2023).
92. Campbell, G. & Tomlinson, A. Transducing the *Dpp* morphogen gradient in the wing of *Drosophila*: regulation of *Dpp* targets by *brinker*. *Cell* **96**, 553–562 (1999).
93. Geetha-Loganathan, P., Nimmagadda, S. & Scaal, M. *Wnt* signaling in limb organogenesis. *Organogenesis* **4**, 109–115 (2008).
94. Abou Azar, F. & Lim, G. E. Metabolic contributions of *wnt* signaling: more than controlling flight. *Front. Cell Dev. Biol.* **9**, 709823 (2021).
95. Lartillot, N. & Poujol, R. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol. Biol. Evol.* **28**, 729–744 (2011).
96. Zdobnov, E. M. et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749 (2017).
97. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
98. Szklarczyk, D. et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).


## Acknowledgements

We thank the National Drosophila Species Stock Center and the Artyom Kopp laboratory for providing fly cultures and technical support; E. Sedore and L. Pekowsky for assistance with Syracuse University's OrangeGrid; B. Woolsey for graphic design assistance; S. Buel for technical assistance; and members of the Center for Reproductive Evolution for intellectual support. This research was funded by the National Science Foundation (grant nos. DEB-1811805 to S.P. and S.D.; DEB-1655840 to S.D. and S.P.; DEB-2011045 to R.A.G.; DEB-2030129, DEB-1839598 and DEB-1241253 to P.M.O.), the National Institutes of Health (grant no. R35GM147454 to Y.H.A.-B,), a Syracuse University Postdoctoral Fellowship (to S.D. and S.P.) and a generous gift from M. and J. Weeden to Syracuse University.

## Author contributions

Conceptualization: S.P., S.D., Z.A.S., D.A.P. Methodology: S.D., S.P., Z.A.S., K.B., R.A.G., B.Y.K. Investigation: Z.A.S., S.P., S.D., R.A.G., K.B., A.A., A.S.C., S.L., B.Y.K., P.M.O., P.W., Y.H.A.-B., C.M.-G. Funding acquisition: S.P., S.D., R.A.G., P.M.O., D.A.P. Project administration: S.P., S.D., P.M.O., B.Y.K. Supervision: S.P., S.D., Z.A.S., D.A.P. Writing—original draft: S.D., S.P., Z.A.S., R.A.G. Writing—review and editing: S.P., S.D., Z.A.S., R.A.G., S.L., B.Y.K., Y.H.A.-B.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41559-024-02587-2.

**Correspondence and requests for materials** should be addressed to Zeeshan A. Syed, Steve Dorus or Scott Pitnick.

**Peer review information** *Nature Ecology and Evolution* thanks Jacek Radwan, Murielle Ålund and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



[1]Center for Reproductive Evolution, Department of Biology, Syracuse University, Syracuse, NY, USA. [2]Department of Entomology, Cornell University, Ithaca, NY, USA. [3]Department of Biology, Stanford University, Stanford, CA, USA. [4]Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA. [5]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland. ✉e-mail: zeeshyeda@gmail.com; sdorus@syr.edu; sspitnic@syr.edu

nature portfolio

Corresponding author(s): Scott Pitnick
Steve Dorus
Zeeshan Syed

Last updated by author(s): Sep 19, 2024

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☐ | ☒ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Morphometric measurements were made using ImageJ (Fiji). The software is not explicitly mentioned in the text since the same data collection method was used in a previous study which is cited in the methodologies (Syed et al 2020) |
| Data analysis | Majority of the analyses were performed using standard functions in R version 4.2.0 onwards. The R packages used in the manuscript are detailed in the manuscript and are listed below:<br>stats v.4.2.0, Phytools v.1.9-16, caper v.1.0.2, OUwie v.2.10, qgg, VCA<br>Additional third-party software used in the manuscript are listed below:<br>BayesTraits v. 2, est-sfs, PLINK, FlyMine, Coevol, OrthoDB, MAFFT, PhyloBayes, GWA as implemented in dgrp2.gnets.ncsu.edu, STRING |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data are available in supplementary files accompanying this paper.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | *Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.*<br>*Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.*<br>*Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.* |
| Reporting on race, ethnicity, or other socially relevant groupings | *Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).*<br>*Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)*<br>*Please provide details about how you controlled for confounding variables in your analyses.* |
| Population characteristics | *Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."* |
| Recruitment | *Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.* |
| Ethics oversight | *Identify the organization(s) that approved the study protocol.* |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | We combined (1) comparative analyses of 149 Drosophila species, (2) a genome-wide association study (GWAS) in D. melanogaster, and (3) molecular evolutionary analysis of ~9400 genes to elucidate how sperm and female sperm-storage organ length coevolved into one of nature's most extreme ornaments and preferences. For the comparative analyses, sperm and SR length were quantified for 149 drosophilid species. The bivariate relationship between these sex-specific traits was examined using Phylogenetic Generalized Least Squares (PGLS) regression, accompanied by examination of pattern and rates of diversification using using state-of-the-art discrete analyses and modeling. All of these comparative analyses were conducted within a phylogenetic framework using a highly robust phylogeny based on full-genome sequencing of all 149 species. Second, the GWAS analyses of this same male ornament (i.e., sperm length) and female preference (i.e., SR length) was conducted for D. melanogaster using 126 lines of the Drosophila Genetic Reference Panel (DGRP). Identification and analyses of single nucleotide polymorphisms (SNPs) significantly associated with among-line variation in the sex-specific traits of interest were conducted using an established DGRP analytical pipeline. Third, the comparative genomic analysis of approximately 9400 orthologous genes was conducted for 15 Drosophila species with annotated genomes using the software Coevol. All subsequent gene analyses (e.g., gene ontology, molecular evolution, selective sweep, network interaction) were performed using standard tools as detailed in the Methods. |

| | |
|---|---|
| Research sample | Comparative phenotypic analyses were conducted on 149 species of drosophilid species and maintained in laboratory culture and obtained from the National Drosophila Species Stock Center and from the laboratories of various colleagues. Species selection was based on availability and to maximize phylogenetic breadth. To the extent possible (some species require unique culture media recipes), all species were cultured under identical rearing conditions. Selection of the 15 species for comparative genomic analyses were based on the availability of sequenced and annotated genomes. |
| Sampling strategy | Sample sizes for quantification of sperm length (ornament size) and female seminal receptacle length (preference strength) among species were based on knowledge of within- and among-species variation in these traits based on numerous previous investigations by the researchers. These sample sizes are shown to be sufficient based on trait variance measures as provided in the manuscript. Sample sizes used to quantify line-specific ornament size and preference strength for the GWAS analysis using 126 DGRP lines were determined by power analyses based on preliminary analyses of much larger sample sizes for both traits in 20 DGRP lines. |
| Data collection | All dissections of sperm from males and preparation of slides for imaging and measurement were performed by Postdoctoral Fellow Zeeshan Syed. All dissection and imaging of female reproductive tracts were performed by Principle Investigator Scott Pitnick. Quantification of sperm and SR length from images for all of the investigations was performed by highly trained undergraduate researchers (who appear as coauthors). |
| Timing and spatial scale | Not applicable. |
| Data exclusions | Data exclusion was limited to the genome-wide association study of sperm length. We Winsorized the outliers in our datasets (based on a Tukey's Outlier test), which adjusted the 'long sperm' outliers and resulted in a phenotype distribution that does not deviate from normality. This correction did not alter any conclusions made in the paper. |
| Reproducibility | Not applicable. |
| Randomization | Randomization procedures were not relevant as no "experiments" were performed. Traits were quantified for populations (149 drosophilid species; 126 isogenic DGRP lines). |
| Blinding | Researchers using the morphometric software FIJI to measure sperm and SR length from micrographs for the GWAS analyses were different from the researchers dissecting the flies and capturing the trait micrographs; they lacked information on the DGRP line identity of micrographs. Otherwise, "blindness" was not relevant to this research program. |

Did the study involve field work? ☐ Yes ☒ No

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☐ | ☒ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

| | |
|---|---|
| Laboratory animals | This study includes laboratory cultures of 149 species of Drosophila and related genera. The text is explicit with regard to species, strains and age of study organisms. |
| Wild animals | No wild animals were used in the reported research. |
| Reporting on sex | This research explicitly investigated phenotypic variation, underlying genetics and evolutionary diversification of sex-specific traits. As such, the text is consistently clear about the sex-specific nature of all data and analyses. |
| Field-collected samples | This research did not include any field-collected samples. |

| | |
|---|---|
| Ethics oversight | All reported research was conducted using members of the insect family Drosophilidae (order Diptera). No ethical approval or guidance was required. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Plants

| | |
|---|---|
| Seed stocks | *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.* |
| Novel plant genotypes | *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.* |
| Authentication | *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.* |